Online Study on Reading Behavior of Data-Rich Texts with Integrated Word-Scale Visualizations

Franziska Huth[*], Lukas Kaminski[†], Fairouz Grioui[*], and Tanja Blascheck[*]
University of Stuttgart

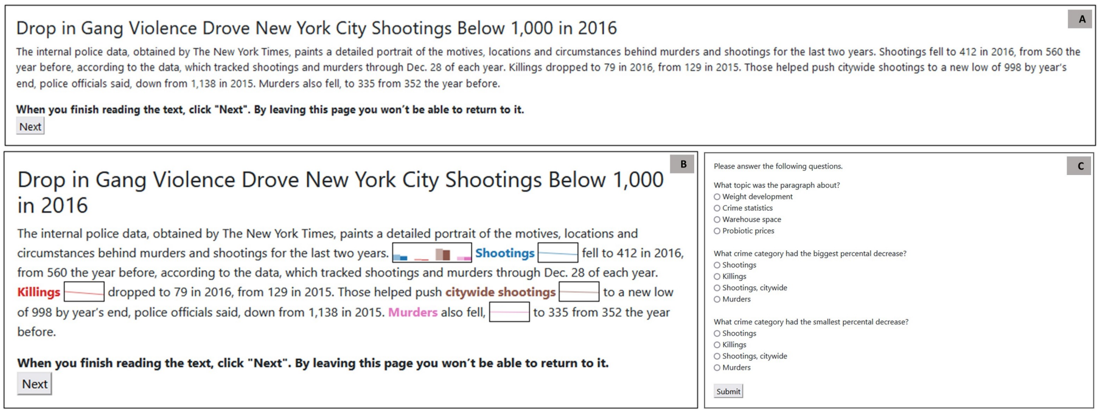


Figure 1: Study stimuli examples (A and B) from task category (i) calculated information. A is the text stimulus in condition 1 (without word-scale visualizations), B is the text stimulus in condition 2 (with word-scale visualizations). C shows the questions we asked after participants read the text (the first question is an attention check).

Abstract

We report on an online study on reading behavior of data-rich texts with integrated word-scale visualizations. We compare task completion time and accuracy of information gain to texts without word-scale visualizations. Although the results do not indicate significant differences for these measures between the two conditions, participants' feedback revealed interesting findings.

Index Terms: Human-centered computing—Visualization—Visualization design and evaluation methods; Human-centered computing—Visualization—Empirical studies in visualization

1 Introduction

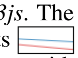
To augment text with information, small graphics that are known as sparklines [16] or word-scale visualizations (*WSV*s) have proven effective in improving understanding of data in a text [3,5,10,11] without distracting readers [15]. Researchers have shown that *WSV*s are useful in settings like explaining eye-tracking data [1,2], augmenting scientific texts [10], better understanding of software code [7], and for increasing discussion topic awareness in social media feeds [8,9]. There are multiple good options where to place *WSV*s in text [6] and how to interact with them [4].

However, there is still little knowledge about how *WSV*s influence reading behavior and whether there are differences between different kinds of visualized data. With an online study, we examined how much faster or slower people read texts with and without *WSV*s, as well as how well they comprehend information from those texts. We chose news articles that contain factual and statistical information, e. g, on health, crime statistics, or finances. We compared four information extraction tasks, and investigated the overall and individual effects on text reading behavior for the conditions with or without *WSV*s embedded, but found no statistically significant differences.

2 Study Design

We chose a between-subjects study design to compare task completion time and answer accuracy on texts without *WSV*s (condition 1) and with *WSV*s (condition 2). To compare performance on different kinds of tasks, we asked questions about the information in the texts in one of four categories: (i) calculated information that could be learned by simple calculations from parameters in the text; (ii) relation of two tendencies of parameter trends in the text; (iii) trend, minimum, and maximum value of a parameter that has a non-monotone development; and (iv) information that is in the text, but not explicitly stated by numbers.

Our hypotheses are that ($H1$) the error rate is lower for condition 2 (with *WSV*s) than for condition 1 (without *WSV*s), ($H2$) task completion time is lower for condition 2 (with *WSV*s) than for condition 1 (without *WSV*s), and ($H3$) there are differences in error rate and completion time between the task types. We presume that without *WSV*s, readers need more mental effort to compute and compare the information in the texts.

After some demographic questions and a training task, participants saw a total of 8 stimuli, 2 in each category, in a randomized order to account for learning effects. For two of the stimuli, we asked an additional question about the text topic as an attention check. We hand-picked the texts for the stimuli from existing news articles from *The New York Times* [12] and used paragraphs containing suitable data for our task types, then embedded *WSV*s using *D3js*. The *WSV*s we used were basic bar charts  and line charts  that we placed directly after the respective textual information, with color-coded words in the texts corresponding to the visual elements. Figure 1 shows example stimuli from category (i), detailed information of all stimuli can be found in our *OSF* repository [13]. At the end of the study, we asked participants about the confidence in the correctness of their answers and the perceived difficulty of the tasks, and for condition 2 whether they focused mostly on the text,

[*]e-mail: [firstname].[lastname]@vis.uni-stuttgart.de
[†]e-mail: lukas.kaminski@fius.stuvus.uni-stuttgart.de

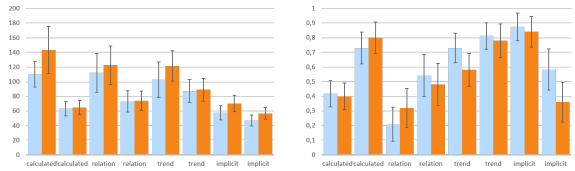


Figure 2: Mean time needed in seconds (left) and rate of correct answers (right) with 95% confidence intervals for each stimulus, comparing the condition without WSVs and with WSVs.

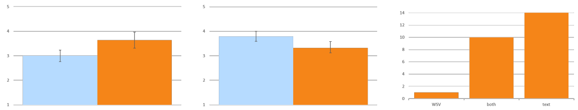


Figure 3: Mean confidence in correctness of participants' given answers, with 95% confidence intervals (left) and perceived task difficulty (middle), on a Likert scale from 1 = low to 5 = high, comparing the condition without WSVs and with WSVs. Number of participants that stated they focused mostly on the visualizations, the text, or both when performing the tasks for condition 2 (right).

the visualizations, or both. We used the online platform *Prolific* [14] to recruit 60 participants who stated they were fluent in English.

3 Study Evaluation

Of 60 participants, we could not use the data of 10 from an initial test run because of technical difficulties with our logs. Further, we excluded one participant for failed attention checks, resulting in usable data from a total of 49 participants, 24 for condition 1 (without WSVs) and 25 for condition 2 (with WSVs). The majority of the participants (41) were between 18 and 40 years old, 8 were between 41 and 64. Twenty seven participants stated they had some experience with visualizations or statistical graphics, 18 said they had no experience, and 4 stated they had a lot of previous experience.

Figure 2 shows task completion time for all stimuli. We find that generally participants took slightly more time in condition 2 (with WSVs), on average 12.35 minutes for completing all stimuli, than in condition 1 (without WSVs), on average 10.86 minutes. The error rate for condition 2 (with WSVs) is slightly lower for 2 out of 8 stimuli and higher for 6 of the stimuli, but not statistically significant. The average rate of correct answers is 0.65 for condition 1 and 0.62 for condition 2. Between some stimuli there are significant differences in time taken and error rate, especially within category (i) with calculated information. However, the differences between category types are not statistically significant.

In Figure 3 we plot the answers to the qualitative questions. The participants in condition 2 (with WSVs) were more confident in the correctness of their answers and rated the tasks easier than the participants in condition 1 (without WSVs), despite not having a lower error rate. Slightly more participants (in condition 2) stated that they focused mostly on text than on the WSVs or both.

4 Discussion

We studied the effects of WSVs on information extraction from data-rich news articles. We compared error rate and task completion time on texts containing different kinds of numeric information visualized with WSVs to texts without WSVs and could not find significant differences between the conditions or between the text categories. Thus, we have to reject all 3 hypotheses.

There was no correlation between participants' performance on the tasks and their age or previous experience with visualizations. In condition 2 (with WSVs) participants who stated that they focused mostly on the visualizations, or both text and visualizations, had a lower error rate compared to those who focused mostly on the text, which might be an interesting starting point for further investigation. In future studies, we would be interested in finding reasons that make some participants rely on WSVs to get the information while others grant them less attention. We also consider exploring eye-tracking techniques to better understand the area of interest (or type of information, i.e. text or WSV) that participants focus on while reading texts and for how long. Finally, a long-term study to investigate the effects of WSVs in data-rich texts on information retention can be considered.

Acknowledgments

We thank all participants of the study. This research was supported by the German Science Foundation (DFG) as part of the Priority Program VA4VGI (SPP 1894) and grant DFG ER 272-14. Tanja Blascheck is funded by the European Social Fund and the Ministry of Science, Research and Arts Baden-Württemberg.

References

[1] F. Beck, Y. Acurana, T. Blascheck, R. Netzel, and D. Weiskopf. An expert evaluation of word-sized visualizations for analyzing eye movement data. In *Workshop on Eye Tracking and Visualization*, ETVIS '16, pp. 50–54. Springer, 2016. doi: 10.1109/ETVIS.2016.7851166

[2] F. Beck, T. Blascheck, T. Ertl, and D. Weiskopf. Word-Sized Eye-Tracking Visualizations. In *Workshop on Eye Tracking and Visualization*, ETVIS '15, pp. 113–128. Springer, 2015. doi: 10.1007/978-3-319-47024-5_7

[3] P. Goffin, J. Boy, W. Willett, and P. Isenberg. An Exploratory Study of Word-Scale Graphics in Data-Rich Text Documents. *IEEE TVCG*, 23(10):2275–2287, 2017. doi: 10.1109/TVCG.2016.2618797

[4] P. Goffin, P. Isenberg, T. Blascheck, and W. Willett. Interaction Techniques for Visual Exploration Using Embedded Word-Scale Visualizations. In *Conference on Human Factors in Computing Systems*, CHI '20, pp. 1–13. ACM, 2020. doi: 10.1145/3313831.3376842

[5] P. Goffin, W. Willett, A. Bezerianos, and P. Isenberg. Exploring the effect of word-scale visualizations on reading behavior. In *Conference on Human Factors in Computing Systems*, CHI '15, pp. 1827–1832. ACM, 2015. doi: 10.1145/2702613.2732778

[6] P. Goffin, W. Willett, J.-D. Fekete, and P. Isenberg. Exploring the placement and design of word-scale visualizations. *IEEE TVCG*, 20(12):2291–2300, 2014. doi: 10.1109/TVCG.2014.2346435

[7] J. Hoffswell, A. Satyanarayan, and J. Heer. Augmenting Code with In Situ Visualizations to Aid Program Understanding. In *Conference on Human Factors in Computing Systems*, CHI '18, pp. 1–12. ACM, 2018. doi: 10.1145/3173574.3174106

[8] F. Huth, M. Awad-Mohammed, J. Knittel, T. Blascheck, and P. Isenberg. Online study of word-sized visualizations in social media. In *EuroVis 2021 - Posters*. The Eurographics Association, 2021. doi: 10.2312/evp.20211069

[9] F. Huth, T. Blascheck, S. Koch, S. Utz, and T. Ertl. Word-sized Visualizations for Exploring Discussion Diversity in Social Media. In *IVAPP 2021-12th International Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pp. 256–265. SCITEPRESS, 2021. doi: 10.5220/0010328602560265

[10] S. Latif and F. Beck. Visually Augmenting Documents With Data. *Computing in Science Engineering*, 20(6):96–103, 2018. doi: 10.1109/MCSE.2018.2875316

[11] B. Lee, N. H. Riche, A. K. Karlson, and S. Carpendale. Sparkclouds: Visualizing trends in tag clouds. *IEEE TVCG*, 16(6):1182–1189, 2010. doi: 10.1109/TVCG.2010.194

[12] The New York Times. https://www.nytimes.com/, 2022.

[13] OSF—Online Study on Reading Behavior of Data-Rich Texts with Integrated Word-Scale Visualizations. https://osf.io/49bhw, 2022.

[14] © 2022 Prolific. https://www.prolific.co/, 2022.

[15] P. Ruchikachorn and P. Rattanawicha. An eye-tracking study on sparklines within textual context. In *EuroVis 2018 - Posters*. The Eurographics Association, 2018. doi: 10.2312/eurp.20181119

[16] E. R. Tufte. *Beautiful Evidence*. Graphics Press, 1 ed., 2006.